https://doi.org/10.2352/ISSN.2169-2629.2017.32.248

Individual Corresponding Colors Data and Chromatic Adaptation Transforms

Shengyan Cai[1,2], Mark D. Fairchild[2]; 1 Packaging & Printing Engineering School, Tianjin University of Science & Technology, Tianjin, China; 2 Munsell Color Science Laboratory, Rochester Institute of Technology, Rochester, New York, United States of America

Abstract

It has been well established, through the study of observer metamerism, that color matching functions vary across individual observers. Chromatic adaptation transforms, however, have not been similarly studied to examine carefully for individual differences. Existing chromatic adaptation transformations, such as CAT02 that is embedded within CIECAM02, are fitted to population data on corresponding colors that tell us little about individual differences and have inherently large uncertainty. It is often stated that fitted von Kries type models fit within the population of observers (e.g. within the inter-observer variability) rather than accurately fitting the mean of the data (e.g. within the standard error of the mean of a population or an individual). That suggests two questions examined in this work: Do such models actually predict the mean results? And can individual data on chromatic adaptation be described with 3x3 von Kries type models. These questions are approached by collecting corresponding colors data with very high accuracy and precision allowing the analysis and examination necessary to begin creating the next generation of accurate chromatic adaptation transforms.

Introduction

Corresponding colors are color stimuli that match across changes in viewing conditions.[1] For example, if a stimulus under incandescent adaptation is matched with another stimulus that appears identical, but is viewed under daylight illumination, then those two stimuli are considered corresponding colors. Currently, chromatic adaptation models for practical applications are loosely based on the von Kries (1902) hypothesis that the separate cone types have independent gain controls with the level set by the stimulation of the cone types themselves and derived by fitting psychophysical data.[2] Prediction errors of these chromatic adaptation transforms, or CATs, are no less than about 4.0 CMC(1:1) or CIE DE2000 units, when evaluating models such as CMCCAT2000, CMCCAT97, CAT02 and CIECAT94.[3] The collected datasets of Luo and Rhodes (1999) that are commonly used in chromatic adaptation modeling are mostly from experiments incorporating only single observations per observer per color and a small number of observers due to the inherent difficulty in the experimental techniques.[4] Kuo et al. (1995) found that typical inter-observer variation for studying chromatic adaptation was about 4 CMC(1:1) units (which are similar in magnitude to CIE DE2000 color difference units).[5] Hence, if a CAT has an error of prediction equal to, or less than, 4 units, it might be considered satisfactory. So the limitation of the number of observers and observations in the experiments is subject to significant improvement to get precise corresponding colors data.

This research firstly aims to collect highly precise individual corresponding colors data. To begin to achieve this objective, a small number of observers and colors, but large number of trials are adopted. Experimental data are evaluated firstly and then utilized to evaluate existent chromatic adaptation models. Finally the necessity of building individual CATs is explored.

Experimental

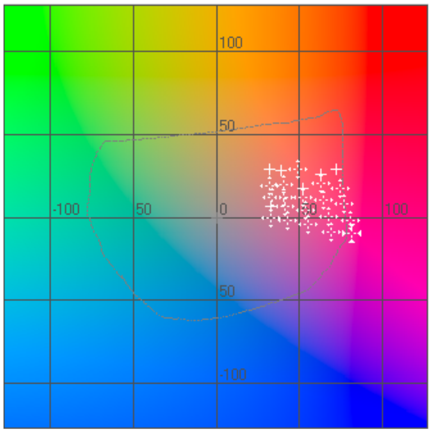
To get highly precise corresponding colors data, a large number of repetitions of the psychophysical experiment is required. In this work, 30 repeat trials are adopted rather than the more typical single trial. Due to the large number of replicates, a small number of colors and observers are used for this stage of the research. Ultimately five colors were evaluated by six observers for 30 trials each across one change in illumination. The experimental stimuli were reflecting objects produced by printing ink on paper. A variety of color patches were provided on reference sheets for observers to choose as matches to the color appearance of a set of five test patches. Observers marked the spatial location of the matching patch on a response sheet with the same layout of color patches as seen on reference sheets. Additionally, to test color appearance memory ability after two minutes of adaptation break, two additional colors were evaluated under illumination identical with the test condition (no change in adaptation).

Preparation of Test Sheets and Reference Sheets

Red, blue, green, yellow and a light (tan) colors were selected for the experiment. It is known from previous research that highly chromatic colors have larger errors in CAT predictions than lower chroma colors. So high chroma red, blue, green and yellow were adopted as test colors referring to the Helson data. Historically, superior datasets are most often reflectance media such as the Lam data,[6] the Helson data, and the Kuo & Luo data. So reflecting media was adopted in this experiment. All the test patches and reference patches were printed on a Canon iPF6400 printer using Onyx Rip-Queue 11 printing software with Gold Fibre Silk paper (310g/m2) and Canon Lucian ink.

The side size of each square color patch was 2.2cm calculated in terms of the viewing distance of about 65cm to produce a 2 degree viewing field. Two test color patches are juxtaposed on each test sheet with 4.5 cm distance between them to avoid direct simultaneous contrast influence on each other. Each test patch had its corresponding reference sheet which has enough patches with enough color range for observers to choose a satisfactory corresponding color. Observation for each test patch was repeated 30 times, the positions of reference patches on reference sheets was varied from trial to trial to avoid being remembered by observers. Thus four sets of 4X4 patch reference sheets were prepared. The four sets of reference sheets had different color patches. To make the four sets of 4X4 patch reference sheets, a set of 5X5 reference sheets was made and then randomly decomposed into four 4X4 sets.

Firstly, the center patch of 5X5 color patches was calculated by CAT02. Then preliminary experiment was implemented to adjust the colors of 5X5 patches to ensure sufficient color range and appropriate increment between adjacent horizontal and vertical patches. At the same time, the colors of all these patches should be within the gamut of the printer. After iterative printing, observation



and adjustment, 5X5 patch reference sheets are determined. Fig. 1 is 5X5 color patches on a*b* plane for the red color center.

Figure 1. An example of the CIELAB sampling used to prepare reference sheets for corresponding colors elections. In this case the red color center is illustrated.

To test the color memory ability of observers, two color patches, blue and yellow, were put on one test sheet. Four sets of corresponding 4X4 patch reference sheets were also made that included colorimetric matches to the test colors. Observers then made memory matches to these two patches following the same experimental protocol but with no change in adaptation.

In one session, five test colors were repeated for two trials. With two color patches on each test sheet, there were five test sheets for the chromatic adaptation experiment. Adding the one memory color test sheets, there are altogether six sheets in an experimental session. The six sheets were printed and spiral bound. The four sets of 12 reference sheets were also printed and spiral bound into four collections denoted with A, B, C and D. In each experimental session each observer was assigned one of the four collections of reference sheets. The spiral binding was minimally obtrusive and a neutral black color.

Preparation of Test Sheets and Reference Sheets

The two adaptation light sources adopted for the experiment were illuminant A and illuminant D65 simulators, which are the most commonly used illumination in previous research. They were provided by incandescent and daylight of GTI Colormatcher CMB-3064 light booth. The mean tristimulus values, in cd/m^2 from a white reference standard, and normalized tristimulus values of the booth's incandescent and daylight illumination are shown in Table 1.

The viewing area of the light booth was 76cm height, 163cm width, and 76cm depth. It was big enough to execute the experiment with room for a wide adapting field of view. When the experiment began, an observer sat in front of the light booth. A sample holder was put in the center of the light booth. The test sheets collection and one reference sheet collection were put respectively on the two sides of the sample holder. Each of six test sheets are required for a round of the experiment. Each round included the four steps outlined below.

Table 1. Absolute (cd/m^2 from a diffuse white reference) and relative tristimulus values of the incandescent (A) and daylight (D65) illumination used in the experiments.

	Incandescent (A)			Daylight (D65)		
	Xrw	Yrw	Zrw	Xw	Yw	Zw
Mean Measurement	473.40	437.60	160.50	325.0	341.20	397.10
Normalized Results	108.18	100.00	36.67	95.25	100.00	116.38

1. The observer takes two minutes to adapt to the incandescent illumination at first.[7]

2. Then the observer would put the test sheet on the sample holder and have 60 seconds to remember the appearance of the two test colors. Observers were allowed to use whatever method they felt would allow most accurate memory of the colors for a short period of time.

3. Then they put the test sheet aside and covered it with a piece of white paper. The light source automatically switched to daylight and the observer again adapted to it for two minutes.

4. After the second adaptation, the observer put the corresponding reference sheet on the sample holder and had 60 seconds to choose the best matches from a 4X4 collection of patches on the reference sheet sets by marking on the answer sheet as shown in Fig. 2. They could mark directly on the location of the matching patch (shaded areas), or select a location between 2 or 4 patches (or even outside all patches) on the grids to indicate the matching color fell between, or outside, the appearance of the patches provided.

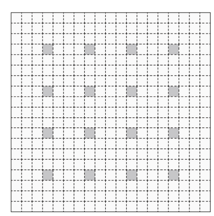


Figure 2. Illustrations of the response sheet that allowed observers to indicate the spatial location of their best matching corresponding color. A different sheet was used for each test color (12 in a session, or set of six rounds).

There were six rounds in total for each replicate trial. The first five rounds were chromatic adaptation experiment, and the sixth round was for the memory test. Each round included two test colors so 12 corresponding colors were measured (two of which were memory tests) in each session. In the memory test experiment, the light source was daylight (D65). During the six round experiment, the light sources switched automatically by programming in advance. An audio recording was made to remind the observers of each step as they proceeded.

Observers performed the experiment 15 times. The four different sets of reference sheets were used in order and circulated through the 15 trials to avoid the possibility of observers learning the spatial locations of responses.

Results & Discussion

The spectral reflectance data of each test patch and reference patch were measured with an X-Rite i1 Pro2. Then their tristimulus values and L*a*b* values were calculated with the spectral data of the two light sources in the viewing booth. The experimental results (L*a*b*) for each observer for 10 reference colors and two memory colors were obtained by interpolation or extrapolation according to the 4X4 color patches of the reference sheets. Their

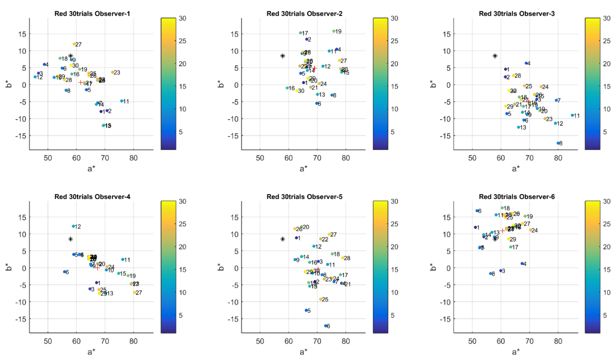


Figure 3. Example results for the red color center illustrating each of the 30 corresponding colors selections for each of six observers.

corresponding tristimulus values were also computed. Since the two groups of five test colors were not completely identical, the two groups of results can't be regarded as two identical trials. The second group of five colors was transformed to be equivalent to the first by CAT02 regarding the two groups of test colors as white points. The two groups of five colors are so close to each other that any error introduced by the transformation can be considered negligible (very small fractions of a CIELAB unit). After the transformation, 30-trial data sets for each of the five test colors were obtained from 15 sessions of experiments. CIELAB a*b* distribution plots for each of the six observers for each color were created. Fig. 3 illustrates the a*b* distribution plot of six observers for the red color center as an example. Results were similar for other colors. Thirty-trial experimental results are denoted with the color filled circles and trial number. The red '+' sign denotes the mean (a*, b*) of 30 trials for the observer and the black '*' sign denotes the CAT02 prediction.

Figure 3. Example results for the red color center illustrating each of the 30 corresponding colors selections for each of six observers.

Reliability Test

The reliability of the experimental results was tested by intra-observer variability and inter-observer variability. Color difference in DE2000 between each trial and mean value and their mean, additional statistics across the 30 trials for each observer and across the six observers were also computed and reviewed. Additionally, the consistency of experimental results was tested by the method proposed by Cai. [8]

Intra-Observer Variability

Color difference in DE2000 for the red color center between each trial and mean value for each observer are shown in Fig. 4. This shows that, on average the CIE DE2000 difference between any given experimental trial and the mean results of 30 trials ranges from a low of about 1 unit up to over 7 units with the average of this variance metric being 2.68 DE2000 units. Taking this metric as a form of standard deviation, except in terms of DE2000, together with the 30 trials, the intra-observer "standard error of the mean" in color difference terms is about 0.5 to 0.65 DE2000 units (probably less than a visual threshold).

The above result is consistent with previous research showing DE2000 values of 4-6 as precision metrics across observers. However, with 30 replicate trials for each observer, the precision with which the mean is known improves by roughly the square root of 30 or about five and half. This increase in precision is just what is needed to test and create improved adaptation models.

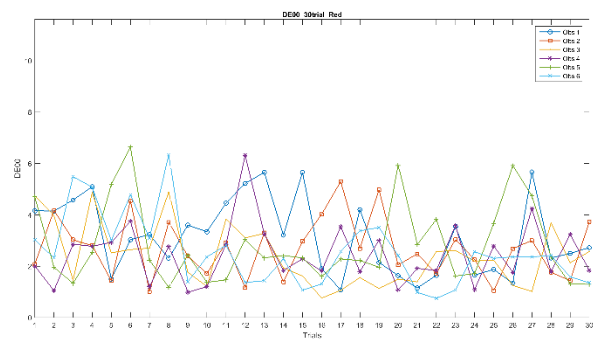


Figure 4. Color difference (DE2000) between each trial and the mean of 30 trials for the red color center for each observer.

Figure 5 illustrates the standard deviation of the observation results across all 30 trials for each observer and for each color center. They are computed in terms of the formula of standard deviation except that (X_i-$\overline{X}$) is replaced by the color difference in DE00 of each trial.

$$sd = \sqrt{\frac{\sum_{i=1}^{n}(X_i\text{-}\bar{X})^2}{n-1}} \qquad (1)$$

These results indicate that the variation in the variation of each individual observer response is quite consistent across observers and across colors. Observers are similar in the magnitude and variability in their uncertainty in selecting colors. As will be seen below this seems to be limited by the accuracy and precision of short term color memory.

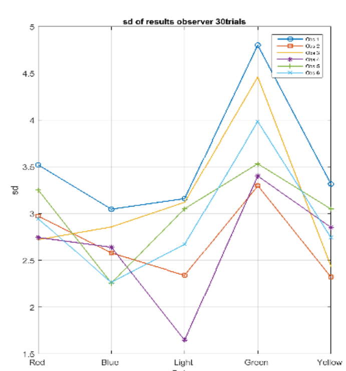


Figure 5. Standard deviations of the observation results from the mean for each observer and color center.

Taking this metric as a form of standard deviation, together with the 30 trials, the intra-observer "standard error of the mean" in color difference terms is about 0.5 to 0.65 DE2000 units (probably less than a visual threshold).

Inter-Observer Variability

The standard deviations of the experimental results across all six observers are also calculated in terms of the similar method as the above and shown in Table 2. These results suggest that observers have different degrees of scatter in different color centers. The light color has the biggest variance across observers (5.26) and followed by green (3.39). The red, blue and yellow colors have similar variance of 2.7. Inter-observer variability shows that there are significantly different results between observers. These differences were confirmed by ANOVA.

Table 2. Standard deviation of the experimental results across all six observers.

Color	Red	Blue	Light	Green	Yellow
sd	2.72	2.94	5.26	3.39	2.61

Consistency Test by ΔE'

Cai et al. proposed a method to test consistency of a corresponding colors dataset.[8] A second transform can be derived by exchanging the two groups of tristimulus values of a corresponding colors dataset. In other words the direction of chromatic adaptation (e.g., A to D65) is reversed (e.g. D65 to A) to see if the same transformation applies. The prediction difference between these two transforms, when applied in the same direction (denoted as ΔE') can illustrate the consistency of the data (under the assumption that an optimized von Kries type model can describe the data).

ΔE' of the experimental data of each observer were calculated and shown in Table 3.

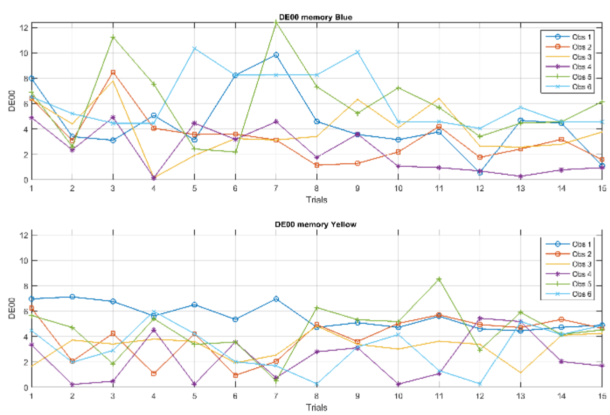


Figure 6. DE2000 between accurate color match (memory) and each trial for six observers and two colors.

Table 3. Consistency evaluation for each of the six observers.

	Observer					
	1	2	3	4	5	6
ΔE'	0.23	0.53	0.55	0.50	0.82	1.14

The ΔE' of the full collection of the experimental data overall is 0.41. Compared with the 8 superior datasets selected in Cai's research (their ΔE' are shown as Table 4), the consistency of the six individual data sets are essentially equivalent with the collection of best previously reported results. This analysis assumes that 3X3 von Kries type models works and should be consistent in each direction.

Table 4. Data set consistency for previous data sets. The current experiment falls well within the best of previous work.

Data	Lam & Rigg	Helson	Kuo & Luo	Lutchi
ΔE'	0.27	0.53	0.7	0.3
Data	Breneman 6	CSAJ	Breneman 1	Kuo & Luo (TL84)
ΔE'	0.66	1.12	1.28	0.67

Validity Test

Validity tests were executed by comparing memory accuracy test with the cross-adaptation corresponding colors results.

Memory Accuracy

Color difference in DE2000 between experimental results of each observer and correct values are shown in Fig. 6 for both the blue and yellow colors tested. The error from the accurate color memory (with a two minute delay for simulated adaptation time) ranges from zero (perfect memory match) up over 12 DE2000 units for the blue color and over eight DE2000 units for the yellow color. This poor accuracy of color memory is consistent with previous work on color memory[9] and supports the theoretical concepts of color categories and so-called color constancy.

Figure 7 illustrates the mean color differences from the mean memory match (the precision of the matches) for each of the six observers and averaged across all observers (right panel) for both colors. These results illustrate that any given trial is most likely to have a DE2000 color difference of just under 2.5 units. This level of precision is comparable with that obtained in adaptation-change conditions suggesting that color memory is the precision-limiting factor in this experimental paradigm.

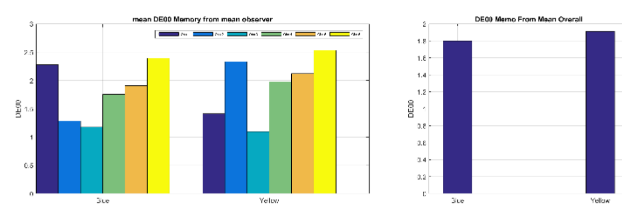


Figure 7. Mean color differences to the mean memory matches for each observer and for all observers.

Figure 8 illustrates the mean color differences from the perfectly accurate memory match (the accuracy of the matches) for each of the six observers and averaged across all observers (right panel) for both colors. These results illustrate that there is a bias in the memory matches. Such a bias will clearly impact chromatic adaptation results and will require further exploration and modeling. It might also suggest that corresponding colors data are asymmetric (e.g. D65-to-A cannot predict directly A-to-D65 matches) which would require a very different type of adaptation model.

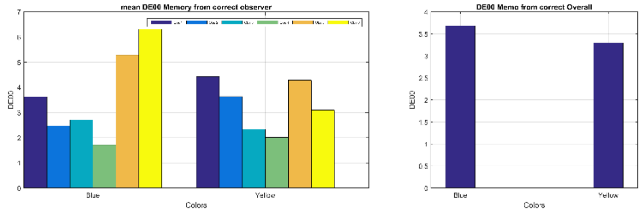


Figure 8. Mean color differences from accurate color memory matches for each observer and averaged across all observers for both colors.

Equality of Memory and Adaptation

Mean color differences (from mean) and standard deviations of results of the memory and adaptation conditions for each observer and each color are shown in Fig. 9. The results show that the mean difference from the mean selected colors is very similar for the memory condition and the adaptation condition for both colors. This test illustrates that the variation in color memory matches is not meaningfully impacted by the change in adaptation state from incandescent to daylight. It helps confirm the validity of this experimental approach for future research.

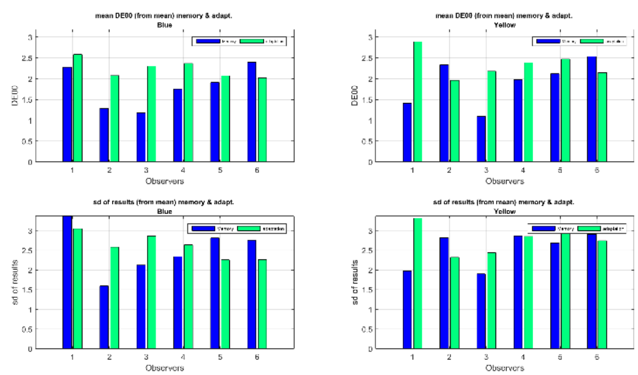


Figure 9. Means and standard deviations of color differences from the mean responses for memory matching and corresponding colors selections.

Performance of Existing CATs on Individual Data

Six CATs were tested using the current experimental results. They are HPE, CAT02, CAT97, CAT2000, T_{Lam} and T_{Com}. T_{Lam} is a sharpening transform derived from the Lam & Rigg data. T_{Com} was proposed by Cai[Error! Bookmark not defined.] with a mixed dataset comprised of four datasets with A/D65 illuminants and reflection media. The six CATs perform similarly with the exception that HPE (a traditional von Kries model) predicts significantly less well for blue and yellow as shown in Fig. 10.

From Fig. 10, one can see that the performances of the six CATs vary in terms of colors. Color differences are about 3.0, 3.5, 8.0, 8.0 and 11.0 in DE2000 for yellow, light, blue, green and red respectively. None of these CATs are capable of predicting the obtained data within the level of uncertainty of the experiment (less than a single DE2000 unit).

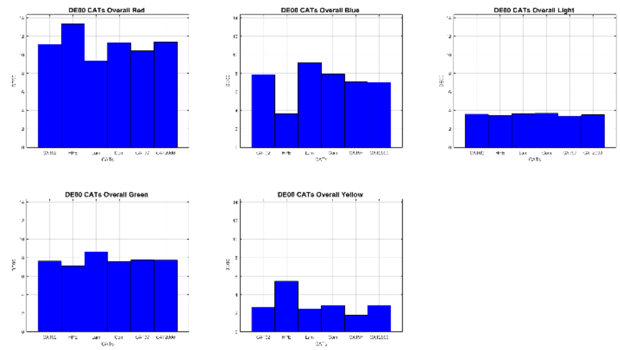


Figure 10. Mean color differences between predictions and the average results of the current experiment for six chromatic adaptation transforms.

With respect to prediction error for each of the six observers, each of the six CATs all perform with a certain systematic tendency. This is illustrated in Fig. 11 for the CAT02 predictions error relative to the six observers.

It is noticeable that CAT02 predictions fall with a systematic bias relative to almost all six observers. This means that CAT02 doesn't perform well in predicting these individual data or the mean data. It can also be clearly seen that there are systematic differences between observers (remember the individual data points are accurate to better than one DE2000 unit).

To test the validity of 3X3 von Kries type models, six CATs are derived from the six individual data and evaluated by these data themselves respectively. If the 3X3 von Kries type model works well, the prediction error should be very small. Table 5 shows the prediction errors of these CATs derived by white-point preserving data-based spectral sharpening method from the six individual data and evaluated by these data themselves.

Table 5. The prediction errors (in DE00) of six CATs derived from six individual data and evaluated by these data themselves respectively.

Observers	1	2	3	4	5	6
Prediction Errors	5.12	5.91	5.92	5.98	6.75	7.26

Table 5 shows that 3X3 von Kries models are not adequate for this precision of the new data by factor of 10. Thus, it is likely necessary to build a new CAT, perhaps a new type of CAT, to adequately fit data with the level of precision that has been obtained.

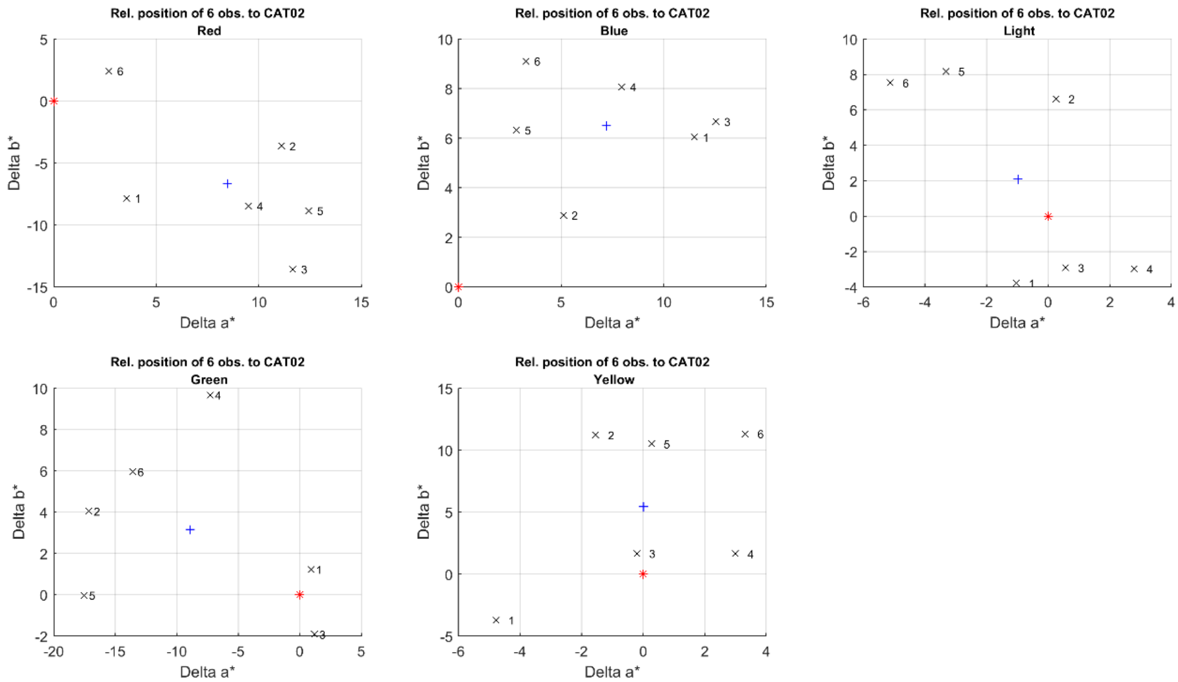


Figure 11. CAT02 predictions (at origin) relative to the mean corresponding colors results for each of six observers (X sign). Red '+' sign denotes the mean results across six observers for each color.*

Conclusions

There are several important conclusions that can be drawn from the current study.

(1) Precise (uncertainty in DE2000 < 1 unit) corresponding colors data can be collected with significant replication.

(2) Individual differences in chromatic adaptation exist.

(3) Existing CATs cannot predict the precise average results or individual differences.

(4) 3X3 von Kries type models do not appear to be adequate to predict the results.

References

[1] Fairchild MD. Color appearance models. John Wiley & Sons Ltd: 2013.

[2] J. von Kries, Chromatic adaptation, Festschrift der Albrecht-Ludwig-Universität, (Fribourg) (1902) [Translation: D.L. MacAdam, Sources of Color Science, MIT Press, Cambridge, (1970)].

[3] CIE. A Review of Chromatic Adaptation Transforms. 2004:29.

[4] M.R. Luo and P.A. Rhodes, Corresponding-color datasets, Color Research and Application 24, 295-296 (1999).

[5] Kuo, W. G., Luo, M. R., & Bez, H. E. (1995). Various chromatic-adaptation transformations tested using new colour appearance data in textiles. Color Research & Application, 20(5), 313-327.

[6] King Man Lam. Metamerism and colour constancy [dissertation]. University of Bradford; 1985.

[7] Shevell S K. The time course of chromatic adaptation[J]. Color Research and Application, 2001, 26: S170-S173.

[8] Shengyan Cai, Qianwen Chen etc. A New Method to Evaluate a Corresponding Color Dataset Based on Its Two Derived Sharpening Transforms. Color Research and Application. Accepted.

[9] Pérez-Carpinell J, Baldoví R, de Fez M D, et al. Color memory matching: Time effect and other factors[J]. Color Research & Application, 1998, 23(4): 234-247.

Author Biographies

Mark Fairchild is Professor and Director of the Program of Color Science and Munsell Color Science Laboratory as well as Associate Dean of Research and Graduate Education for the College of Science at Rochester Institute of Technology.

Shengyan Cai is a visiting scholar at RIT's Munsell Color Science Laboratory and a faculty member in the Packaging & Printing Engineering School of the Tianjin University of Science and Technology.